Weekly Report

2018.0709-2018.0715

1.This Week

Power Grid Deep Learning Paper
1.We keep on training with a 15K dataset. The model performance hasn't yet reach our expectation and we are adjusting the parameters which may takes a long time. (Sometimes the models overfits so we try to add regularization options and dropout layers, and adjust the learning rate, batch size as most works suggests to solve this problem.)
2.We compute the effectiveness of each training sample by computing the shanon entropy of it. According to the computed shannon entropy, we sample buses from each instance to imitate incomplete datasets and try to use a small set of this data set to train the model (to see this is a possible thing to do). The result is depressing for now, the accurary is about 30% on both training set, validation set and test set (no matter how we adjust our parameter). We'll check the data next week to see if there contains to many random info or random pattern in the dataset that results in this situation.

SQC Paper
1. Zongzhuang is realizing the CUSUM algorithm with the real case power grid data (data of 36 nodes in the power grid). I asked him to update his work and share with us, but the ddl is passed and I haven't received any thing yet.

Others
1.Read machine learning blogs about how to solve overfitting problems.
2.revise the waveline paper.
3.Write the paper tricks (of system papers) following the writing structures and contributions we organized last week. The plan:

时间	计划进度	完成情况
7.2-7.9	1.确定任务与分工 2.明确整理思路	已完成
7.10-7.29	1.完成各自分工的套路整理	进展中
7.30-8.8	1.汇总各自整理的内容 2.组织讨论并互相补充	
8.9-8.22	1.根据讨论补充相应内容	

8.23-8.31	1.二次汇总	

WaveLines Revision:
changes to be made:
【done】1.make clear the definitions of power gris simulation terms like transient simulation stc..(make a form)
2.clearly define all patterns mentioned in this paper and discuss how existing works distinguish these patterns in the related work.
3.revise the evaluation part: add quantitative evaluation of how accurate wavelines can help to find patterns(effenctiveness), how long will take(efficiency) and complete understanding (comprehensiveness).
【done】4.explain the data preprocessing into a whole continuous process.
【done】5.related work: explain why methods used to visualize pairwise variables mentioned are not used in this paper.(use this to guide the alternative designs discussion.)
【done】6.use less words but add a figure to explain the section waveline design trade-off.
【done】7.explain more carefully why we use a bulb metaphor.

Paper Reading
1.Multiclass-Multilabel Learning when the Label Set Grows with the Number of Examples.
2.Label Embedding Trees for Large Multi-Class Tasks
3.Fast and Balanced: Efficient Label Tree Learning for Large Scale Object Recognition

These three papers target at solving the same problem: improve the accuracy of multiclass classification with large number of labels (classes). And these three papers are listed here by there publishing time. The first one tries to remove labels the classifier performs poorly on. And the second one wants to improve the performance as well as accuracy by learning a structure over the set of classes (a label tree). It learns the label tree by optimizing the overall tree loss and embeds labels in a low dimensional space to improve the efficiency. Motivated by the second paper, Li Feifei and her team proposed the third paper. They advances the label tree technique by introducing a way to simultaneously determine the structure of the tree and learn the classifiers for each node in the tree. They also allows fine grained control over the efficiency vs accuracy trade-off in designing a label tree, leading to a more balanced tree. (Experiment validates that it performs well with 10184 classes and 9 million images.)

I'm reading label tree papers and its theory because it is useful for us when we have a large number of classes (and we will). I try to find an open source code but failed. So if we want to try it, we need to have a full understanding of its theory and code by ourselves.

2.Progress

Work	Deadline	Progress
Power grid paper with Deeping learning	-	1.Adjust the model parameter. 2.Generating data.
SQC Paper	-	About to started.
WaveLine revision	ASAP	Follow the revision plan to revise the paper.